NOVEL MUTATIONS IN HUMAN MLH1 AND HUMAN MSH2 GENES USEFUL IN DIAGNOSING COLORECTAL CANCER

INTRODUCTION

This application claims the benefit of U.S. provisional
5   application Serial No. 60/105,180, filed October 22, 1998.

BACKGROUND OF THE INVENTION

Colorectal cancer (CRC) is one of the most common fatal
cancers in developed countries, and the worldwide incidence
is increasing. The United States and the United Kingdom are
10 high incidence countries, with an estimated 133,500 new cases
and 55,300 deaths (Parker et al. CA Cancer J. Clin. 1996 46:5-
27) in the United States and 30,941 cases and approximately
17,000 deaths in the United Kingdom (HMSO UK Cancer Registry
Data). The population lifetime risk is 1 in 25 in the United
15 States and Northern Europe and thus represents a significant
public   health   issue   (Sharp   et   al.   Cancer   Registration
Statistics Scotland 1981-1990, Information and Statistics
Division, The National Health Service in Scotland, Edinburgh
(1993)). Identification of people who are predisposed to the
20 disease   would   allow   targeting   of   effective   preventative
measures with the aim of reducing the considerable cancer
related mortality (Burke et al. J. Am. Med. Ass'n. 1997
227:915-919).

One group of people with a very high colorectal cancer
25 risk are those who carry germline mutations in genes that
participate in DNA mismatch repair. hMSH2 (Fishel et al. Cell
1993 75:1027-1038; Leach et al. Cell 1993 75:1215-1225; U.S.
Patent 5,591,826) and hMLH1 (Bronner et al. Nature 1994
368:258-261; Papadopoulos et al. Science 1994 263:1625-1629;

PCT Publication No. WO 95/20678, published on August 3, 1995)
are the two genes most commonly involved in heredity
predisposition to CRC, but mutations in hPMS1 and hPMS2 also
occur in a minority of cases (Nicolaides et al. Nature 1994

5   371:75-80). Such mutations are usually associated with marked
familial aggregation of colorectal, uterine and other cancers
constituting the clinically defined autosomal dominant
syndrome of hereditary non-polyposis colorectal cancer (HNPCC)
(Lynch et al. Gastroenterology 1993 104:1535-1549; Liu et al.

10 Nature Med. 1996 2:169-174; Wijnen et al. Am. J. Hum. Genet.
1995 56:1060-1066; Mary et al. Hum. Mol. Genet. 1994 3:2067-
2069; Nystrom-Lahti et al. Nature Med. 1995 1:1203-1206).
However, an appreciable proportion of patients who have early
onset colorectal cancer but who do not fulfill pragmatic

15 criteria for HNPCC (Vasen et al. Dis. Colon Rectum 1991
34:424-425) also carry mismatch repair gene mutations (Liu et
al. Nature Med. 1995 2:169-174; Dunlop et al. Br. Med. J. 1997
314:1779-1780). Thus, restricting genetic testing to
individuals from families fulfilling HNPCC criteria is likely

20 to exclude a significant fraction of gene carriers in the
general population. However, screening unselected patients
with sporadic cancer represents an enormous workload and may
provide a very low yield of mutation carriers (Liu et al. Nat.
Med. 1995 1:348-352; Tomlinson et al. J. Med. Genet. 1997

25 34:39-42).

It is clear that issues concerning indications for
genetic testing and interpretation of results are critical in
hereditary cancer syndromes (Giardiello et al. N. Engl. J.
Med. 1997 336: 823-827).

30   Using a population-based approach, factors indicative
of the likelihood of identifying patients with mismatch repair
gene mutations were investigated. Improved approaches to
mutation detection and the prevalence of detectable mismatch
repair gene alterations in various screened groups who were

not selected on the basis of family history were also determined.

SUMMARY OF THE INVENTION

An object of the present invention is to provide novel, variant hMLH1 sequences.

Another object of the present invention is to provide novel, variant hMSH2 sequences.

Another object of the present invention is to provide a method of diagnosing hereditary non-polyposis colorectal cancer in a patient or determining a patient's susceptibility to developing hereditary non-polyposis colorectal cancer via detection of novel variant hMLH1 or hMSH2 sequences or the exonic or intronic sequences of the hMLH1 and hMSH2 genes.

Another object of this invention is to provide methods and compositions for identifying new variants of hMLH1 and hMSH2 genes.

Yet another object of the present invention is to provide experimental models of hereditary non-polyposis colorectal cancer.

DETAILED DESCRIPTION OF THE INVENTION

To better elucidate the structure of human MLH1 and human MSH2 genes and to determine possible sites of alternative splicing, the genes were cloned and sequenced and PCR was used to determine alternate splice products (variants) and exon/intron boundaries. Elucidation of intron/exon boundary sequences revealed that hMLH1 is encoded by 19 coding exons. The hMLH1 gene sequence was determined by PCR.

The intron/exon structure of the hMLH1 is shown below. Positions of introns that interrupt the hMLH1 cDNA are shown. Exonic sequence is presented in upper case and intronic sequence in lower case letters. Exons are numbered from the 5' end of the cDNA sequence.

hMLH1 Exon 1

aggcactgaggtgattggc (SEQ ID NO:1)

tgaaggcacttccgttgagcatctagacgtttccttggctcttctggcgccaaa (SEQ ID NO:2)

5 ATGTCGTTCGTGGCAGGGGTTATTCGGCGGCTGGACGAGACAGTGGTGAACCGCATCGC GGCGGGGGAAGTTATCCAGCGGCCAGCTAATGCTATCAAAGAGATGATTGAGAACTG (SEQ ID NO:3)

gtacggagggagtcgagccgg (SEQ ID NO:4)

gctcacttaagggctacga (SEQ ID NO:5)

10 cttaacgg (SEQ ID NO:6)

hMLH1 Exon 2

aatatgtacattagagtagttg (SEQ ID NO:7)

cagactgataaattattttctgtttgatttgccag (SEQ ID NO:8)

TTTAGATGCAAAATCCACAAGTATTCAAGTGATTGTTAAAGAGGGAGGCCTGAAGTTGA

15 TTCAGATCCAAGACAATGGCACCGGGATCAGG (SEQ ID NO:9)

gtaagtaaaacctcaaagtagcaggatgtttgtgcgcttcatggaa (SEQ ID NO:10)

gagtcaggacctttctctg (SEQ ID NO:11)

hMLH1 Exon 3

agagatttggaaaatgagtaac (SEQ ID NO:12)

20 atgattatttactcatcttttggtatctaacag (SEQ ID NO:13)

AAAGAAGATCTGGATATTGTATGTGAAAGGTTCACTACTAGTAAACTGCAGTC CTTTGAGGATTTAGCCAGTATTTCTACCTATGGCTTTCGAGGTGAG (SEQ ID NO:14)

gtaagctaaagattcaagaaatgtgtaaaatat (SEQ ID NO:15)

cctcctgtgatgacattgt (SEQ ID NO:16)

25 c

hMLH1 Exon 4

aacctttccctttggtgagg (SEQ ID NO:17)

tgacagtgggtgacccagcagtgagttttctttcagtctattttcttttcttccttag (SEQ ID NO:18)

30 GCTTTGGCCAGCATAAGCCATGTGGCTCATGTTACTATTACAACGAAAACAGCT GATGGAAAGTGTGCATACAG (SEQ ID NO:19)

gtatagtgctgacttctttactcatatatattcattctgaaatgtattttgg (SEQ ID
NO:20)

gcctaggtctcagagtaatc (SEQ ID NO:21)

hMLH1 Exon 5

5   ttgatat (SEQ ID NO: 22)

gattttctcttttcccttggg (SEQ ID NO:23)

attagtatctatctctctactggatattaatttgttatattttctcattag (SEQ ID
NO: 24)

AGCAAGTTACTCAGATGGAAAACTGAAAGCCCCTCCTAAACCATGTGCTGGCAATCAAG

10 GGACCCAGATCACG (SEQ ID NO: 25)

gtaagaatggtacatgggaca (SEQ ID NO:26)

gtaaattgttgaagctttgtttg (SEQ ID NO:27)

hMLH1 Exon 6

gggttttattttcaagtacttctatg (SEQ ID NO: 28)

15 aatttacaagaaaaatcaatcttctgttcag (SEQ ID NO: 29)

GTGGAGGACCTTTTTTACAACATAGCCACGAGGAGAAAAGCTTTAAAAAATCCAAGT

GAAGAATATGGGAAAATTTTGGAAGTTGTTGGCAG (SEQ ID NO:30)

gtacagtccaaaatctgggagtgggtctctgagatttgtcatcaaagtaatgtgttctagt

(SEQ ID NO:31)

20 gctcatacattgaacagttgctgagc (SEQ ID NO:32)

hMLH1 Exon 7

ctagtgtgtgttttggc (SEQ ID NO:33)

aactcttttcttactcttttgttttcttttccag (SEQ ID NO:34)

GTATTCAGTACACAATGCAGGCATTAGTTTCTCAGTTAAAAAA (SEQ ID NO:35)

25 gtaagttcttggtttatgggggatggttttgtttttatgaaaagaaaaaggggatttttt

aatagtttgct (SEQ ID NO:36)

ggtggagataaggttatg (SEQ ID NO:37)

hMLH1 Exon 8

ctcagccatgagacaataaatcc (SEQ ID NO:38)

30 ttgtgtcttctgctgtttgtttatcag (SEQ ID NO:39)

CAAGGAGAGACAGTAGCTGATGTTAGGACACTACCCAATGCCTCAACCGTGGACAATAT
TCGCTCCATCTTTGGAAATGCTGTTAGTCG (SEQ ID NO:40)
gtatgtcgataacctatataaaaaaatcttttacatttattatcttggtttatcatt (SEQ ID NO:41)
ccatcacattatttgggaacc (SEQ ID NO: 42)

hMLH1 Exon 9

caaaagcttcagaatctc (SEQ ID NO: 43)
ttttctaatag (SEQ ID NO:44)
AGAACTGATAGAAATTGGATGTGAGGATAAAACCCTAGCCTTCAAAATGAATGGTTACA
TATCCAATGCAAACTACTCAGTGAAGAAGTGCATCTTCTTACTCTTCATCAACC (SEQ ID NO:45)
gtaagttaaaaagaaccacatgggaaat (SEQ ID NO:46)
ccactcacaggaaacacccacag (SEQ ID NO:47)

hMLH1 Exon 10

catgactttgtgtgaatgtacacc (SEQ ID NO: 48)
tgtgacctcacccctcaggacagttttgaactggttgctttctttttattgtttag (SEQ ID NO:49)
ATCGTCTGGTAGAATCAACTTCCTTGAGAAAAGCCATAGAAACAGTGTATGCAGCCTATT
TGCCCAAAAACACACACCCATTCCTGTACCTCAG (SEQ ID NO:50)
gtaatgtagcaccaaactcctcaaccaagactcacaaggaa (SEQ ID NO:51)
cagatgttctatcaggctctcctc (SEQ ID NO: 52)

hMLH1 Exon 11

gggctttttctcccctccc (SEQ ID NO:53)
actatctaaggtaattgttctctcttattttcctgacag (SEQ ID NO: 54)
TTTAGAAATCAGTCCCCAGAATGTGGATGTTAATGTGCACCCCACAAAGCATGAAG
TTCACTTCCTGCACGAGGAGAGCATCCTGGAGCGGGTGCAGCAGCACATCGAGAGCAAG
CTCCTGGGCTCCAATTCCTCCAGGATGTACTTCACCCAG (SEQ ID NO:55)
gtcagggcgcttctcatccagctacttctctggggcctttgaaatgtgcccggccaga (SEQ ID NO:56)
cgtgagagcccagatttt (SEQ ID NO:57)

hMLH1 Exon 12

aattatacctcatactagc (SEQ ID NO:58)

ttctttcttagtactgctccatttggggacctgtatatctatacttcttattctgagtct

ctccactatatatatatatatatattttttttttttttttttttaatacag (SEQ

5 ID NO:59)

ACTTTGCTACCAGGACTTGCTGGCCCCTCTGGGGAGATGGTTAAATCCACAACAAGTCT

GACCTCGTCTTCTACTTCTGGAAGTAGTGATAAGGTCTATGCCCACCAGATGGTTCGTA

CAGATTCCCGGGAACAGAAGCTTGATGCATTTCTGCAGCCTCTGAGCAAACCCC

TGTCCAGTCAGCCCCAGGCCATTGTCACAGAGGATAAGACAGATATTTCTAGTGGCAGGG

10 CTAGGCAGCAAGATGAGGAGATGCTTGAACTCCCAGCCCCTGCTGAAGTGGCTGCCAAAA

ATCAGAGCTTGGAGGGGGATACAACAAAGGGGACTTCAGAAATGTCAGAGAAGAGAGGAC

CTACTTCCAGCAACCCCAG (SEQ ID NO:60)

gtatggccttttgggaaaagtacagccta (SEQ ID NO:61)

cctcctttattctgtaataaaac (SEQ ID NO:62)

15 **hMLH1 Exon 13**

tgcaacccacaaaatttggc (SEQ ID NO:63)

taagtttaaaaacaagaataataatgatctgcacttccttttcttcattgcag (SEQ ID

NO:64)

AAAGAGACATCGGGAAGATTCTGATGTGGAAATGGTGGAAGATGATTCCCGAAAGGAAA

20 TGACTGCAGCTTGTACCCCCCGGAGAAGGATCATTAACCTCACTAGTGTTTTGAGTCTCCAG

GAAGAAATTAATGAGCAGGGACATGAGG (SEQ ID NO:65)

gtacgtaaacgctgtggcctgcctgggatgcatagggcctcaactgccaa (SEQ ID NO:

66)

ggttttggaaatggagaaag (SEQ ID NO:67)

25 **hMLH1 Exon 14**

tggtgtctctagttctgg (SEQ ID NO: 68)

tgcctggtgctttggtcaatgaagtggggttggtaggattctattacttacctgttttt

tggttttatttttttgttttgcag (SEQ ID NO:69)

TTCTCCGGGAGATGTTGCATAACCACTCCTTCGTGGGCTGTGTGAATCCTCAGTGGGCCTTG

30 GCACAGCATCAAACCAAGTTATACCTTCTCAACACCACCAAGCTTAG (SEQ ID NO:70)

gtaaatcagctgagtgtgtgaacaa (SEQ ID NO:71)

gcagagctactacaacaatg (SEQ ID NO: 72)

hMLH1 Exon 15

cccatttgtcccaactgg SEQ ID NO:73

ttgtatctcaagcatgaattcagcttttccttaaagtcacttcattttattttcag (SEQ ID NO:74)

5 TGAAGAACTGTTCTACCAGATACTCATTTATGATTTTGCCAATTTTGGTGTTC TCAGGTTATCG (SEQ ID NO:75)

gtaagtttagatccttttcactt (SEQ ID NO:76)

ctgacatttcaactgaccg (SEQ ID NO:77)

hMLH1 Exon 16

10 catttggatgctccgttaaagc (SEQ ID NO:78)

ttgctccttcatgttcttgcttcttcctag (SEQ ID NO:79)

GAGCCAGCACCGCTCTTTGACCTTGCCATGCTTGCCTTAGATAGTCCAGAGAGTGGCTG GACAGAGGAAGATGGTCCCAAAGAAGGACTTGCTGAATACATTGTTGAGTTTCTGAAGA AGAAGGCTGAGATGCTTGCAGACTATTTCTCTTTGGAAATTGATGAG (SEQ ID NO:80)

15 gtgtgacagccattcttatacttctgttgtattctc (SEQ ID NO:81)

caaataaaatttccagccgggtg (SEQ ID NO:82)

hMLH1 Exon 17

ggaaaggcactggagaaatggg (SEQ ID NO:83)

atttgtttaaactatgacagcattatttcttgttcccttgtccttttttcctgcaagcag (SEQ ID NO:84)

20 GAAGGGAACCTGATTGGATTACCCCTTCTGATTGACAACTATGTGCCCCCTTTGGAGGG ACTGCCTATCTTCATTCTTCGACTAGCCACTGAG (SEQ ID NO:85)

gtcagtgatcaagcagatactaagcattt (SEQ ID NO:86)

cggtacatgcatgtgtgctggaggg (SEQ ID NO:87)

25 **hMLH1 Exon 18**

taagtagtctgtgatctccg (SEQ ID NO:88)

tttagaatgagaatgtttaaattcgtacctattttgaggtattgaatttctttggaccag (SEQ ID NO:89)

GTGAATTGGGACGAAGAAAAGGAATGTTTTGAAAGCCTCAGTAAAGAATGCGCTATGTT

30 CTATTCCATCCGGAAGCAGTACATATCTGAGGAGTCGACCCTCTCAGGCCAGCAG (SEQ ID NO:90)

gtacagtggtgatgcacactggcaccccaggacta (SEQ ID NO:91)

ggacaggacctcatacat (SEQ ID NO:92)

hMLH1 Exon 19

gacaccagtgtatgttgg (SEQ ID NO:93)

5 gatgcaaacagggaggcttatgacatctaatgtgttttccag (SEQ ID NO:94)

AGTGAAGTGCCTGGCTCCATTCCAAACTCCTGGAAGTGGACTGTGGAACACATTGTC
TATAAAGCCTTGCGCTCACACATTCTGCCTCCTAAACATTTCACAGAAGATGGAAATATC
CTGCAGCTTGCTAACCTGCCTGATCTATACAAAGTCTTTGAGAGGTGTTAA (SEQ ID
NO:95)

10 atatggttatttatgcactgt (SEQ ID NO:96)

gggatgtgttcttctttctc (SEQ ID NO:97)

tgtattccgatacaaagtgttgtatcaaagtgtgatatacaaagtgtaccaacataagtg
(SEQ ID NO:98)

        Elucidation of intron/exon boundary sequences revealed
15 that hMSH2 is encoded by 16 coding exons. The hMSH2 gene
sequence was determined by PCR.

        The intron/exon structure of the hMSH2 is shown below.
Positions of introns that interrupt the hMSH2 cDNA are shown.
Exonic sequence is presented in upper case and intronic
20 sequence in lower case letters. Exons are numbered from the
5' end of the cDNA sequence.

hMSH2 Exon 1

ggcgggaaacagcttagtgggtgtggggtcg (SEQ ID NO:99)

cgcattttcttcaaccagga (SEQ ID NO:100)

25 ggtgaggaggtttcgac (SEQ ID NO:101)

ATGGCGGTGCAGCCGAAGGAGACGCTGCAGTTGGAGAGCGCGGCCGAGGTCGGCTTCGTG
CGCTTCTTTCAGGGCATGCCGGAGAAGCCGACCACCACAGTGCGCCTTTTCGACCGGGG
CGACTTCTATACGGCGCACGGCGAGGACGCGCTGCTGGCCGCCCGGGAGGTGTTCAAGA
CCCAGGGGGTGATCAAGTACATGGGGCCGGCAG (SEQ ID NO:102)

30 gtgagggccgggac (SEQ ID NO:103)

ggcgcgtgctggggagg (SEQ ID NO:104)

gac

hMSH2 Exon 2

gaa

gtccagctaatacagtgcttg (SEQ ID NO:105)

aacatgtaatatctcaaatctgtaatgtactttttttttttttaag (SEQ ID NO:106)

5 GAGCAAAGAATCTGCAGAGTGTTGTGCTTAGTAAAATGAATTTTGAATCTTTTGTAAAA

GATCTTCTTCTGGTTCGTCAGTATAGAGTTGAAGTTTATAAGAATAGAGCTGGAAATAAG

GCATCCAAGGAGAATGATTGGTATTTGGCATATAAG (SEQ ID NO:107)

gtaattatcttcctttttaatttacttattttt (SEQ ID NO:108)

ttaagagtagaaaaataaaaatgtg (SEQ ID NO:109)

10 aag

hMSH2 Exon 3

ATTAATAAGGtTCATAGAGTTTGGATTTTTCCtTTTtgc (SEQ ID NO:110)

ttataaaattttaaagtatgttcaag (SEQ ID NO:111)

agtttgttaaatttttaaaattttatttttacttag (SEQ ID NO:112)

15 GCTTCTCCTGGCAATCTCTCTCAGTTTGAAGACATTCTCTTTGGTAACAATGAT

ATGTCAGCTTCCATTGGTGTTGTGGGTGTTAAAATGTCCGCAGTTGATGGCCAGAGACAG

GTTGGAGTTGGGTATGTGGATTCCATACAGAGGAAACTAGGACTGTGTGAATTCCCTGAT

AATGATCAGTTCTCCAATCTTGAGGCTCTCCTCATCCAGATTGGACCAAAGGAATGTGT

TTTACCCGGAGGAGAGACTGCTGGAGACATGGGGAAACTGAGACAG (SEQ ID NO:113)

20 gtaagcaaattgagtctagtgat (SEQ ID NO:114)

agaggagattccaggcctaggaaag (SEQ ID NO:115)

gc

TCTTTAATTGACATGATACTG (SEQ ID NO:116)

hMSH2 Exon 4

25 ttca

tttttgcttttcttattccttttc (SEQ ID NO:117)

tcatagtagtttaaactatttctttcaaaatag (SEQ ID NO:118)

ATAATTCAAAGAGGAGGAATTCTGATCACAGAAAGAAAAAAAGCTGACTTTTCCACAAA

AGACATTTATCAGGACCTCAACCGGTTGTTGAAAGGCAAAAAGGGAGAGCAGATGAATA

30 GTGCTGTATTGCCAGAAATGGAGAATCAG (SEQ ID NO:119)

gtacatggattataaatgtgaattacaatatatataatgtaaatatgtaatatataata

aataatatgtaaactatagtgacttt (SEQ ID NO:120)

ttagaaggatatttctgtca (SEQ ID NO:121)

tat

5   **hMSH2 Exon 5**

actggcacca (SEQ ID NO:122)

gtggtatagaaatcttcgattttt (SEQ ID NO:123)

aaattcttaattttag (SEQ ID NO:124)

GTTGCAGTTTCATCACTGTCTGCGGTAATCAAGTTTTTAGAACTCTTATCAGATGATTC

10 CAACTTTGGACAGTTTGAACTGACTACTTTTGACTTCAGCCAGTATATGAAATTGGATA

TTGCAGCAGTCAGAGCCCTTAACCTTTTTCAG (SEQ ID NO:125)

gtaaaaaaaaaaaaaaaaaaaa (SEQ ID NO:126)

aaaagggttaaaaatgttgatt (SEQ ID NO:127)

gg

15 TTAAAAAATGTTT (SEQ ID NO:128)

t

caTTGACATATACTGAAGAAGCT (SEQ ID NO:129)

TATAAAGGAGCTAAAATATTTGGAAAT (SEQ ID NO:130)

att

20 ATTATACTTGGATTAGATAACTAGCTTTAAATGGGTGTATTTT (SEQ ID NO:131)

    hMSH2 Exon 6

gtt

ttcactaatgagcttgccattc (SEQ ID NO:132)

tttctatttтattttttgtttactag (SEQ ID NO:133)

25 GGTTCTGTTGAAGATACCACTGGCTCTCAGTCTCTGGCTGCCTTGCTGAATAAGTGTAA

AACCCCTCAAGGACAAAGACTTGTTAACCAGTGGATTAAGCAGCCTCTCATGGATAAGA

ACAGAATAGAGGAGAG (SEQ ID NO:134)

gtatgttattagtttatactttcgttagttttatgtaacctgca (SEQ ID NO:135)

gttacccacatgattatacc (SEQ ID NO:136)

30 ac

hMSH2 Exon 7

ga

cttacgtgcttagttgataa (SEQ ID NO:137)

attttaattttatactaaaatattttacattaattcaagttaatttatttcag (SEQ ID
5 NO:138)

ATTGAATTTAGTGGAAGCTTTTGTAGAAGATGCAGAATTGAGGCAGACTTTACAAGAAG
ATTTACTTCGTCGATTCCCAGATCTTAACCGACTTGCCAAGAAGTTTCAAAGACAAGCA
GCAAACTTACAAGATTGTTACCGACTCTATCAGGGTATAAATCAACTACCTAATGTTAT
ACAGGCTCTGGAAAAACATGAAG (SEQ ID NO:139)

10 gtaacaagtgattttgtttttttg (SEQ ID NO:140)

ttttccttcaactcatacaatata (SEQ ID NO:141)

tac

hMSH2 Exon 8

ga

15 tttgtattctgtaaaatgagatcttt (SEQ ID NO:142)

ttatttgtttgtttttactacttctttttag (SEQ ID NO:143)

GAAAACACCAGAAATTATTGTTGGCAGTTTTTGTGACTCCTCTTACTGATCTTCGTTCT
GACTTCTCCAAGTTTCAGGAAATGATAGAAACAACTTTAGATATGGATCAG (SEQ ID
NO:144)

20 gtatgcaatatactttttaatttaag (SEQ ID NO:145)

cagtagttattttttaaaaagcaaag (SEQ ID NO:146)

gcc

hMSH2 Exon 9

gt

25 ctttacccattatttataggatt (SEQ ID NO:147)

ttgtcactttgttctgtttgcag (SEQ ID NO:148)

GTGGAAAACCATGAATTCCTTGTAAAACCTTCATTTGATCCTAATCTCAGTGAA
TTAAGAGAAATAATGAATGACTTGGAAAAGAAGATGCAGTCAACATTAATAAGTGCAGC
CAGAGATCTTG (SEQ ID NO:149)

30 gtaagaatgggtcattggag (SEQ ID NO:150)

gttggaataattcttttgtctat (SEQ ID NO:151)

ac

hMSH2 Exon 10

gg

tagtaggtatttatggaatactttt (SEQ ID NO:152)

tcttttcttcttgtttatcaag (SEQ ID NO:153)

5   GCTTGGACCCTGGCAAACAGATTAAACTGGATTCCAGTGCACAGTTTGGATATTACTTTC

GTGTAACCTGTAAGGAAGAAAAAGTCCTTCGTAACAATAAAAACTTTAGTACTGTAGATA

TCCAGAAGAATGGTGTTAAATTTACCAACAG (SEQ ID NO:154)

gtttgtaagtcattattatatttttaaccctttatt (SEQ ID NO:155)

aattccctaaatgctctaaca (SEQ ID NO:156)

10   tg

hMSH2 Exon 11

ca

cattgcttctagtacacattt (SEQ ID NO:157)

taatatttttaataaaactgttatttcgatttgcag (SEQ ID NO:158)

15   CAAATTGACTTCTTTAAATGAAGAGTATACCAAAAATAAAACAGAATATGAAGAAGCCC

AGGATGCCATTGTTAAAGAAATTGTCAATATTTCTTCAG (SEQ ID NO:159)

gtaaacttaatagaactaa (SEQ ID NO:160)

taatgttctgaatgtcacctg (SEQ ID NO:161)

g

20   **hMSH2 Exon 12**

at

tcagtattcctgtgtacattt (SEQ ID NO:162)

tctgtttttattttttatacag (SEQ ID NO:163)

GCTATGTAGAACCAATGCAGACACTCAATGATGTGTTAGCTCAGCTAGATGCTGTTGTC

25   AGCTTTGCTCACGTGTCAAATGGAGCACCTGTTCCATATGTACGACCAGCCATTTTGGAGAA

AGGACAAGGAAGAATTATATTAAAAGCATCCAGGCATGCTTGTGTTGAAGTTCAAGATG

AAATTGCATTTATTCCTAATGACGTATACTTTGAAAAAGATAAACAGATGTTCCACATC

ATTACTG (SEQ ID NO:164)

gtaaaaaacctggttt (SEQ ID NO:165)

30   ttgggctttgtggggggtaa (SEQ ID NO:166)

cg

hMSH2 Exon 13

cg

cgattaatcatcagtgtac (SEQ ID NO:167)

agtttaggactaacaatccatttattagtagcagaaagaagtttaaaatcttgctttct

5   gatataatttgtttttgtag (SEQ ID NO:168)

GCCCCAATATGGGAGGTAAATCAACATATATTCGACAAACTGGGGTGATAGTACT

CATGGCCCAAATTGGGTGTTTTGTGCCATGTGAGTCAGCAGAAGTGTCCATTGTGGACTG

CATCTTAGCCCGAGTAGGGGCTGGTGACAGTCAATTGAAAGGAGTCTCCACGTTCATGGC

TGAAATGTTGGAAACTGCTTCTATCCTCAG (SEQ ID NO:169)

10   gtaagtgcatctcctagtccctt (SEQ ID NO:170)

gaagatagaaatgtatgtctctg (SEQ ID NO:171)

tcc

hMSH2 Exon 14

ta

15   ccacattttatgtgatgggaa (SEQ ID NO:172)

atttcatgtaattatgtgcttcag (SEQ ID NO:173)

GTCTGCAACCAAAGATTCATTAATAATCATAGATGAATTGGGAAGAGGAACTTCTACCTA

CGATGGATTTGGGTTAGCATGGGCTATATCAGAATACATTGCAACAAAGATTGGTGCTTT

TTGCATGTTTGCAACCCATTTTCATGAACTTACTGCCTTGGCCAATCAGATACCAACTGT

20   TAATAATCTACATGTCACAGCACTCACCACTGAAGAGACCTTAACTATGCTTTATCAGGT

GAAGAAAG (SEQ ID NO:174)

gtatgtactattggagtactctaaattcagaacttg

gtaatgggaaacttactacc (SEQ ID NO:175)

cc

25   **hMSH2 Exon 15**

ct

cttctcatgctgtcccctc (SEQ ID NO:176)

acgcttccccaaatttcttatag (SEQ ID NO:177)

GTGTCTGTGATCAAAGTTTTGGGATTCATGTTGCAGAGCTTGCTAATTTCCCTAAGCAT

30   GTAATAGAGTGTGCTAAACAGAAAGCCCTGGAACTTGAGGAGTTTCAGTATATTGGAGA

ATCGCAAGGATATGATATCATGGAACCAGCAGCAAAGAAGTGCTATCTGGAAAGAGAG

(SEQ ID NO:178)

gtttgtcagtttgtttt (SEQ ID NO:179)

catagtttaacttagcttctc (SEQ ID NO:180)

tat

hMSH2 Exon 16

ta

5   attactcatgggacattcaca (SEQ ID NO:181)

tgtgtttcag (SEQ ID NO:182)

CAAGGTGAAAAAATTATTCAGGAGTTCCTGTCCAAGGTGAAACAAATGCCCTTTAC

TGAAATGTCAGAAGAAACATCACAATAAAGTTAAAACAGCTAAAAGCTGAAGTAATAGC

AAAGAATAATAGCTTTGTAAATGAAATCATTTCACGAATAAAAGTTACTACGTGA (SEQ

10   ID NO:183)

aaa

atcccagtaatggaatgaag (SEQ ID NO:184)

gta

        hMLH1 and hMSH2 genes were sequenced in 50 cancer

15   patients (age of onset <30) and 26 random anonymous donors.
     Initial genomic sequencing detected 12 germline mutations in
     12 patients (24%). Five mutations were found in hMLH1, and
     7 in hMSH2. Using a combination of genomic sequencing and *in
     vitro* synthesized-protein-truncation assay (IVSP), a total of

20   15 germ-line mutations were identified. The mutations are
     described in Table 1.

     **Table 1:   Pathogenic hMLH1 and hMSH2 Mutations Identified in
     Young Colorectal Cancer Probands**

Gene and Patient	Mutation	Nucleotide Change	Effect on Coding Sequence	Location
hMLH1:				
329	616delAAG	Deletion of AAG at 1846-1848	Deletion of Lys616	Exon 16
533	IVS8-3delTA	Deletion of TA at 677-3	Splice mutation	IVS 8

696	K618A	AA→GC at 1852-1853	Lys→Ala at 618	Exon 16
804	R659X	C→T at 1975	Arg→Stop at 659	Exon 17
815	IVS1+1G→A	G→A at 116+1	Splice mutation	IVS 1
817	del exon 13	Deletion of ~3 kb involving IVS 12 through exon 13 to IVS 13	Deletion of codons 470-520 (exon 13)	IVS 12-13, exon 13
889		not identified	Truncation of IVSP	Exons 12-19
hMSH2:				
528	R406X	C→T at 1216	Arg→Stop at 406	Exon 7
579	H639Y IVS13-1G→T	C→T at 1915 G→T at 2211	Double mutation results in deletion of codons 588-820 (exons 12-14)	Exon 12, IVS 13
814	Q601X	C→T at 1801	Gln→Stop at 601	Exon 12
818	Q252X	C→T at 754	Gln→Stop at 252	Exon 4
825	delCTGT	Deletion of CTGT at 808-811	Deletion of codons 265-314 (exon 5)	Exon 5
830	R680X	C→T at 2038	Arg→Stop at 680	Exon 13
1157	M1L	A→T at 1	New initiation at codon 26	Exon 1

"IVS" means intervening sequence.

Two of the mutations identified in Table 1 for hMLH1 and three of the mutations identified in Table 1 for hMSH2 are believed

to be new. For hMLH1, these include: the splice mutation IVS1 + 1G→A in patient 815, also referred to herein as "hMLH1 mutant 1"; and deletion of exon 13 in patient 817, also referred to herein as "hMLH1 mutant 2". For hMSH2, these

5 include the double mutation H639Y IVS13-1G→T leading to deletion of codons 588-820 in patient 579, also referred to herein as "hMSH2 mutant 1", mutation R680X in patient 830 which comprises a nucleotide change from C to T at position 2038 in Exon 13 and results in a stop codon at position 680

10 of the coding sequence, also referred to herein as "hMSH2 mutant 2"; and mutation M1L in patient 1157 which comprises a nucleotide change from A to T at position 1 resulting in a new initiation at codon 26, also referred to herein as "hMSH2 mutant 3". Detection of these genetic mutations is useful in

15 diagnosing HNPCC in a patient and determining susceptibility of a patient for developing HNPCC.

         There are several methodologies available from recombinant DNA technology which may be used for detecting these new variants and identifying additional genetic

20 mutations responsible for colon cancer. The identification of intronic sequences of hMLH1 and hMSH2 provided herein is particularly useful for design of intronic such as those exemplified in SEQ ID NO:1, 5, 7, 11, 12, 16, 17, 21, 23, 27, 28, 32, 33, 37, 38, 42, 43, 47, 48, 52, 53, 57, 58, 62, 63,

25 67, 68, 72, 73, 77, 78, 82, 83, 87, 88, 92, 93, 97, 100, 104, 105, 109, 111, 115, 117, 121, 123, 121, 123, 127, 129, 132, 136, 137, 141, 142, 146, 147, 151, 152, 156, 157, 161, 162, 166, 167, 171, 172, 175, 176, 180, 181 and 184 for use in identifying mutants in the splice donor or acceptor sites of

30 the hMLH1 or hMSH2 gene. Examples of methodologies useful in detecting and identifying new variants of these genes include, but are not limited to, direct probing, ligase chain reaction (LCR) and polymerase chain reaction (PCR) methodology.

         Detection of variants or mutants using direct probing

35 involves the use of oligonucleotide probes which may be

prepared synthetically or by nick translation. In a preferred embodiment, the probes are complementary to at least a portion of the variant hMLH1 or hMSH2 genes identified herein. The DNA probes may be suitably labeled using, for example, a

5 radiolabel, enzyme label, fluorescent label, or biotin-avidin label, for subsequent visualization in for example a Southern blot hybridization procedure. The labeled probe is reacted with a sample of DNA from a patients suspected of having HNPCC bound to nitrocellulose or Nylon 66 substrate. The areas that

10 carry DNA sequences complementary to the labeled DNA probe become labeled themselves as a consequence of the reannealing reaction. The areas of the filter that exhibit such labeling may then be visualized, for example, by autoradiography.

Alternative probe techniques, such as ligase chain

15 reaction (LCR) involve the use of a mismatch probe, *i.e.*, probes which have full complementarity with the target except at the point of the mutation or variation. The target sequence is then allowed to hybridize both with the oligonucleotides having full complementarity, *i.e.*,

20 oligonucleotides complementary to the hMLH1 or hMSH2 variants of the present invention, and oligonucleotides containing a mismatch under conditions which will distinguish between the two. By manipulating the reaction conditions, it is possible to obtain hybridization only where there is full

25 complementarity. If a mismatch is present, then there is significantly reduced hybridization.

The polymerase chain reaction (PCR) is a technique that amplifies specific DNA sequences. Repeated cycles of denaturation, primer annealing and extension carried out with

30 a heat stable enzyme Taq polymerase leads to exponential increases in the concentration of desired DNA sequences.

Given the knowledge of nucleotide sequences encoding the hMLH1 and hMSH2 genes, it is possible to prepare synthetic oligonucleotides complementary to the sequences which flank

35 the DNA of interest. Each oligonucleotide is complementary

to one of the two strands. The DNA is then denatured at high temperatures (e.g., 95°C) and then reannealed in the presence of a large molar excess of oligonucleotides. The oligonucleotides, oriented with their 3' ends pointing towards

5   each other, hybridize to opposite strands of the target sequence and prime enzymatic extension along the nucleic acid template in the presence of the four deoxyribonucleotide triphosphates. The end product is then denatured again for another cycle. After this three-step cycle has been repeated

10 several times, amplification of a DNA segment by more than one million fold can be achieved. The resulting DNA may then be directly sequenced in order to locate any genetic alterations. Alternatively, the identified hMLH1 and hMSH2 variants of the present invention make it possible to prepare oligonucleotides

15 that will only bind to altered DNA, so that PCR will only result in the multiplication of the DNA if the mutation is present. Following PCR, allele-specific oligonucleotide hybridization may be used to detect the colon cancer point mutation.

20 Alternatively, an adaptation of PCR called amplification of specific alleles (PASA) can be employed; this method uses differential amplification for rapid and reliable distinction between alleles that differ at a single base pair. Newton et al. Nucleic Acid Res. 1989 17:2503; Nichols et al. Genomics

25 1989 5:535; Okayama et al. J. Lab. Clin. Med. 1989 1214:105; Sarkar et al. Anal. Biochem. 1990 186:64; Sommer et al. Mayo Clin. Proc. 1989 64:1361; Wu, Proc. Nat'l Acad. Sci. USA 1989 86:2757; and Dutton et al. Biotechniques 1991 11:700. PASA involves amplification with two oligonucleotide primers such

30 that one is allele specific. The desired allele is efficiently amplified, while the other allele(s) is poorly amplified because it mismatches with a base at or near the 3' end of the allele specific primer. Thus, PASA or the related method PAMSA can be used to specifically amplify one or more

35 mutant hMLH1 or hMSH2 alleles. Where such amplification is

performed on genetic material obtained from a patient, it can serve as a method of detecting the presence of one or more mutant hMLH1 and/or hMSH2 alleles in a patient. PCR-induced mutation restriction analysis, often referred to as IMRA, can
5 also be used in the detection of mutants.

Also important is the development of experimental models of HNPCC. Such models can be used to screen for agents that alter the degenerative course of HNPCC. Having identified specific mutations in the hMLH1 and hMSH2 genes as a cause of
10 HNPCC, it is possible using genetic manipulation, to develop transgenic model systems and/or whole cell systems containing a mutated hMLH1 and/or hMSH2 gene or a portion thereof. The model systems can be used for screening drugs and evaluating the efficacy of drugs in treating HNPCC. In addition, these
15 model systems provide a tool for defining the underlying biochemistry of hMLH1 and hMSH2 and their relationship to HNPCC, thereby providing a basis for rational drug design.

One type of cell system which can be used in the present invention can be naturally derived. For this, blood samples
20 from an affected individual are obtained and permanently transformed into a lymphoblastoid cell line using, for example, Epstein-Barr virus. Once established, such cell lines can be grown continuously in suspension cultures and can be used in a variety of *in vitro* experiments to study hMLH1
25 and hMSH2 expression and processing. Another cell line used in these studies comprises skin fibroblasts derived from patients.

The mutated gene can also be excised for use in the creation of transgenic animals containing the mutated gene.
30 For example, the hMLH1 and hMSH2 variants of the present invention can each be cloned and placed in a cloning vector. Examples of cloning vectors which can be used include, but are not limited to, lCharon35, cosmid, or yeast artificial chromosome. The variant hMLH1 or hMSH2 gene can then be
35 transferred to a host nonhuman knockout animal such as a

knockout mouse. As a result of the transfer, the resultant transgenic nonhuman animal will preferably express one or more of the variant hMLH1 or hMSH2 polypeptides.

Alternatively, minigenes encoding variant hMLH1 or hMSH2
5 polypeptides can be designed. Such minigenes may contain a cDNA sequence encoding a variant hMLH1 or hMSH2 polypeptide, preferably full-length, a combination of hMLH1 or hMSH2 exons, or a combination thereof, linked to a downstream polyadenylation signal sequence and an upstream promoter (and
10 preferably enhancer). Such a minigene construct will, when introduced into an appropriate transgenic host, such as a mouse or rat, express a variant hMLH1 or hMSH2 polypeptide.

One approach to creating transgenic animals is to target a mutation to the desired gene by homologous recombination in
15 an embryonic stem (ES) cell *in vitro* followed by microinjection of the modified ES cell line into a host blastocyst and subsequent incubation in a foster mother. Frohman et al. Cell 1989 56:145. Alternatively, the technique of microinjection of the mutated gene, or portion thereof,
20 into a one-cell embryo followed by incubation in a foster mother can be used. Additional methods for producing transgenic animals are well known in the art.

Transgenic animals are used in the assessment of new therapeutic compositions and in carcinogenicity testing, as
25 exemplified by U.S. Patent 5,223,610. These animals are also used in the development of predictive animal models for human disease states, as exemplified in U.S. Patent 5,221,778. Therefore, the novel mutations of the hMLH1 and hMSH2 genes of the present invention, which are believed to cause HNPCC,
30 provide a useful means for developing knockout transgenic animals to assess this disease.

Site directed mutagenesis and/or gene conversion can also be used to a mutate a non human hMLH1 or hMSH2 gene allele, either endogenously or via transfection, such that the

mutated gene encodes a polypeptide with an altered amino acid as described in the present invention.

In addition, antibodies to the hMLH1 or hMSH2 gene and variants thereof can be raised for use in the examination of
5   the function of the truncated transcripts of the hMLH1 or hMSH2 gene. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library.
10 Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the hMLH1 and hMSH2 genes of the present invention can be obtained by direct injection into an animal or by administering the gene to an animal,
15 preferably a nonhuman. The antibody so obtained will then bind the hMLH1 or hMSH2 gene or itself. In this manner, even a fragment of the gene can be used to generate thee antibodies.

For preparation of monoclonal antibodies, any technique
20 which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler et al. Nature 1975 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al. Immunology Today 1983 4:72), and the EBV-hybridoma technique
25 to produce human monoclonal antibodies (Cole et al. in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., 1985, pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce
30 single chain antibodies to the hMLH1 or hMSH2 genes of this invention. Also, transgenic mice may be used to express humanized antibodies to the hMLH1 or hMSH2 genes of this invention.

The following nonlimiting examples are provided to
35 further illustrate the present invention.

EXAMPLES

Example 1: Patients and Samples

A total of 76 subjects were studied: 50 unrelated patients diagnosed with colorectal cancer at <30 years of age

5 and 26 anonymous donors. There were 15 male and 11 female anonymous donors who were cancer free at the time of sampling and whose mean ages was 41 years. None of the study subjects were referred specifically because of a family history of colon cancer. All cancer patients had histologically

10 confirmed colorectal cancer.

Peripheral blood was drawn from each subject and DNA was purified from peripheral-blood leukocytes.

Example 2: Genomic Sequencing

15 DNA was extracted from peripheral blood using the Nucleon DNA Extraction Kit, Scotlab, Lanarkshire, U.K. or using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN) as per the manufacturer's instructions. Each exon of hMSH2 and hMLH1 was amplified by PCR using 40

20 ng of genomic DNA in a volume of 50 μL. Final reaction concentrations were 1 x PCR Buffer II (Perkin Elmer), 3.0 mM $MgCl_2$ (or 1.5 mM for hMSH2 exon 1), 0.2 mM dNTPs, 10 pmol of each specific oligonucleotide primer, and 1.25 units of Taq polymerase. Amplification was hot-started at 94°C for 3

25 minutes, followed by 35 cycles of 94°C for 20 seconds; 55°C for 20 seconds; 72°C for 40 seconds. The final reaction was extended at 72°C for 10 minutes, followed by storage at 4°C. Cycle sequencing used the PRISM Ready Dye Terminator Cycle Sequencing kit with AmpliTaq DNA polymerase, FS (Taq-FS;

30 Perkin Elmer/Applied Biosystems) and an Applied Biosystems DNA Sequencer model 373A or 377 (Parker et al. BioTechniques 1996 21:694-699) according to the manufacturer's instructions. DNA sequence analysis was performed using Sequencer 3.0 (Gene Codes, Ann Arbor, MI)

35 software by comparing published genomic sequences of hMLH1

(Han et al. Hum. Mol. Genet. 1995 4:237-242; Kolodner et al. Cancer Res. 1995 55:242-248) and hMSH2 (Kolodner, et al. Genomics 1994 24:516-526) with that of cancer cases or of random donors.

5 Examples of primers used for mutations in patients 815, 830 and 1157 are as follows:

(1) Patient 815, splice error in hMLH1 exon 1:

Forward primer:

5'-TGTAAAACGACGGCCAGTCTGAGGTGATTGGCTGAAG-3' (SEQ ID NO:

10 185)

Reverse primer:

5'-GGAAACAGCTATGACCATGCCGTTAAGTCGTAGCCCTT-3' (SEQ ID NO:

186)

(2) Patient 830, premature stop codon in hMSH2 exon

15 13:

Forward primer:

5'-TGTAAAACGACGGCCAGTCGATTAATCATCAGTGTAC-3' (SEQ ID NO:

187)

Reverse primer:

20 5'-GGAAACAGCTATGACCATGCAGAGACATACATTTCTATCTTC-3'(SEQ ID NO:

188)

(3) Patient 1157, missense in initial ATG of hMSH2 (exon 1):

Forward primer:

25 5'-TGTAAAACGACGGCCAGTCGCATTTTCTTCAACCAGGA-3'(SEQ ID NO:

189)

Reverse primer:

5'-GGAAACAGCTATGACCATGCCTCCCCAGCACGCGCC-3' (SEQ ID NO: 190)

Example 2: *In Vitro* **Synthesized-Protein-Truncation**

30 **Assay (IVSP)**

cDNA was generated by reverse transcription of RNA purified from lymphoblastoid cell lines from the affected

index case. PCR amplification of the cDNA was used to
introduce a 17-bp consensus T7 promoter sequence and a
mammalian translation-initiation sequence in frame with a
unique hMLH1 or hMSH2 sequence. PCR primer sequences and

5 conditions were similar to those previously described in
Example 1. Each gene was amplified in two or three
overlapping segments. Resultant PCR products were used in
a coupled transcription-translation reaction (Promega)
incorporating 2-5 μCi of ^{35}S-methionine. Labeled in vitro-

10 transcribed protein products from the reaction were heat
denatured and were analyzed by use of 8%, 10% and 12% SDS-
PAGE gels. Gels were washed in fixative and were
autoradiographed overnight at room temperature. All
samples showing truncated protein products were reamplified

15 independently, and an additional IVSP analysis was
performed for conformation. For each analysis, normal
control samples were run in parallel, and the wild-type
full length protein was noted. In most analyses,
artifactual bands were visible, presumably due to

20 internally initiations since these were visible in samples
form normal controls.

Example 3:   Long Range PCR

For long range PCR of the novel mutation of hMSH2
discovered in patient 817, the GeneAmp XL PCR Kit (Perkin

25 Elmer) was used with the following primers:
Forward primer:
5'-GGCCATTGTCACAGAGGATAAGA-3' (SEQ ID NO: 191)
Reverse primer:
5'-ACACAGCCCACGAAGGAGTG-3' (SEQ ID NO: 192)

30 The reaction mixture contained about 400 ng of genomic DNA in
a volume of 50 μL. Final reaction concentrations were 1 x PCR
Buffer II (Perkin Elmer), 1.5 mM Mg(OAc)$_2$, 0.8 mM dNTPs, 40
pmol of each specific oligonucleotide primer, and 4 units of
rTth DNA polymerase. Amplification was hot-started at 94°C

for 1 minute, followed by 26 cycles of 94°C for 15 second and 68°C for 10 minutes. The final reaction was extended at 72°C for 10 minutes, followed by storage at 4°C.

Replicate cDNA sequencing of samples from patient 817
5 reproducibly demonstrated a truncation in hMSH2 due to deletion of the entire exon 13. However, extensive genomic sequencing failed to identify the mutation at the DNA level. Hence, the intronic region around exon 13 was analyzed by long range PCR to determine whether any large genomic deletion had
10 completely removed that exon. Forward primer was in exon 12 and reverse in exon 14, giving around 15.5 kb wild type product. Using this approach, patient 817 was shown to carry a large deletion of approximately 3 kb which resulted in removal of exon 13.

15 **Example 4: Characterization of mutation in patient 579**

Characterization of the mutation in patient 579 was more complex. Replicate hMSH2 IVSPs for patient 579 detected a very short protein fragment, which could not be explained on the basis of the His→Tyr mutation at codon 639, identified by
20 genomic sequencing. Accordingly, additional genomic sequencing needed to be performed which resulted in identification of the second mutation at the splice acceptor site of exon 14. Using restriction-site changes induced by each mutation, both variants were traced through the family
25 and were shown to reside on the same allele. Extensive sequencing of the reverse transcription-PCR products revealed that this complex double mutation results in an in-frame deletion of exons 12-14, thus accounting for the very short IVSP fragment. A His→Tyr mutation at codon 639 which results
30 in a surrogate splice donor site and a 92-bp frameshift deletion of nucleotides 1914-2006, generating a premature termination codon 17 bp downstream of the exon 13 splice acceptor site has been described previously by Leach et al. Cell 1993 75:1215-1225 and Liu et al. Cancer Res. 1994

54:4590-4594. However, the 92 bp splice mutation reported to be present in this mutation was not present in patient 579, thus confirming that the double mutation in patient 579 is distinct from that reported by Liu et al. Cancer Res. 1994
5   54:4590-4594.

All publications including, but not limited to, patents and patent applications, cited in this specification, are herein incorporated by reference as if each individual publication were specifically and individually indicated to
10 be incorporated by reference herein as though fully set forth.

The above description fully discloses the invention, including preferred embodiments thereof. Modifications and improvements of the embodiments specifically disclosed herein are within the scope of the following claims. Without further
15 elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. Therefore, the examples provided herein are to be construed as merely illustrative and are not a limitation of the scope of the present invention in any way.
20 The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows.